Abstract Title Page
Not included in page count.

Title: A Bayesian Perspective on Methodologies for Drawing Causal Inferences in Experimental and Non-Experimental Settings.

Author(s): David Kaplan
Department of Educational Psychology
University of Wisconsin – Madison

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

In recent years, attention in the education community has focused on the need for evidenced-based research, particularly educational policies and interventions that rest on “scientifically based research". The emphasis on scientifically based research in education has led to a corresponding increase in studies designed to provide strong warrants for causal claims. These studies have been conducted relying almost entirely on analytic procedures that are rooted in the frequentist (classical) paradigm of statistics. Over the last decade, however, significant advances have been made in the area of Bayesian statistical inference, owing mostly to computational developments and readily available software (see e.g. Gilks, Richardson, & Spiegelhalter, 1996). With these advances have come important applications of Bayesian methods to problems in the social and behavioral sciences. However, few applications of Bayesian methods to educational problems can be found, with the exception of Bayesian statistical methods for item response theory models in educational measurement (e.g. Fox & Glas, 2001). This talk proposes that Bayesian methods offer an alternative analytic paradigm that can support and enhance scientifically based research in the educational sciences.

The importance of examining statistical modeling in the educational sciences from a Bayesian perspective cannot be overestimated. For too long, statistical methods applied to educational problems have rested on frequentist statistical hypothesis testing, originally developed by Fisher (1941/1925), and then later by Neyman & Pearson (1928). These approaches have been criticized as logically incoherent and that the Neyman-Pearson approach to hypothesis testing in particular has possibly done considerable damage to progress in the social and behavioral sciences. For interesting discussions on this problem, see Harlow, Mulaik, and Steiger (1994)

An internally consistent and coherent alternative to the Neyman-Pearson paradigm in statistics lies with the Bayesian school. The Bayesian alternative to statistical inference provides a rational approach to incorporating uncertainty in statistical models. For example, the frequentist perspective holds that parameters are fixed, and estimates of these fixed parameters are obtained from sample data. The Bayesian perspective, on the other hand, holds that because parameters are unknown, they are subject to the laws of probability, and sensible probability models can be formed to describe their behavior. For important treatments on the elements on Bayesian theory see e.g. Box and Tiao (1973) and Gelman, Carlin, Stern, and Rubin (2003).

In the case of statistical models, the frequentist perspective does not acknowledge that models themselves are sampled from a larger universe of possible models, none of which are true in any sense of the word. The goal of statistical inference from the Bayesian perspective, therefore, is to ascertain which model is favored by the data (Hoeting, Madigan, Raftery, & Volinsky, 1999; Kass and Raftery, 1995). Finally, albeit not a criticism of frequentist statistics per se, it is not uncommon to read papers with descriptions of, say, confidence intervals, that are not correctly interpreted from a frequentist point of view, but are correctly interpreted from a Bayesian point of view. This suggests that the Bayesian perspective aligns with common sense notions of hypothesis testing.

Purpose / objective / research question / focus of study:
Description of what the research focused on and why.

The purpose of this talk is to present an argument for the Bayesian alternative to empirical research in the educational sciences. Empirical research in the educational sciences consists of a common set of designs and analytic strategies where the Bayesian perspective should provide new and important insights. These settings include (a) randomized experimental designs, (b) quasi-experimental/observational designs, and (c) longitudinal studies. Because of time constraints, this talk will focus on Bayesian hypothesis testing in the context of randomized designs using Bayesian ANOVA with "informative hypotheses" (Hoijtink, Klugkist, & Boelen, 2008). We will also to introduce the notion of the Bayesian propensity score (McCandless, Gustafson, & Austin, 2009). The talk will close with a brief discussion of the "elicitation" of priors (Garthwaite, Kadane, & O'Hagan, 2005)

Setting:
Description of where the research took place.

NA

Population / Participants / Subjects:
Description of participants in the study: who (or what) how many, key features (or characteristics).

NA

Intervention / Program / Practice:
Description of the intervention, program or practice, including details of administration and duration.

NA

Research Design:
Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

This talk represents an "analytic essay" that will argue for the Bayesian alternative to frequentist statistics as a means improving quantitative practice in the educational sciences.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

NA

Findings / Results:
Description of main findings with specific details.

NA

Conclusions:
Description of conclusions and recommendations based on findings and overall study.

The significance of this talk is three-fold. First, when reviewing the extant literature, we find that Bayesian statistical methods have not been fully examined as an alternative modeling strategy in the educational sciences. Nevertheless, theoretical and computational developments now offer Bayesian statistical procedures to study designs commonly applied to empirical educational research, including randomized experimental designs, quasi-experimental designs, and longitudinal designs. Thus, because empirical research in the educational sciences utilizes all of these design possibilities, it is wise to examine alternative statistical models that might provide deeper insights into substantively important educational problems.

Second, the Bayesian perspective directly challenges the conventional approach to hypothesis testing in the educational sciences. The Bayesian perspective forces investigators to shift their focus away from conventional null hypothesis testing and many of the ad hoc behaviors that emerge from that approach, and toward closer examination of substantively important parameters and the range of values they can take.

Third, this analytic essay will introduce the notion elicitation as a means of formally incorporating prior information from either empirical meta-analytic studies or expert opinion. Although addressing the problem of elicitation is not the main thrust of this proposal, the potential benefits are significant. Specifically, research on the elicitation of priors can provide guidance on how to effectively integrate content area knowledge with statistical theory so that expert groups and key educational stakeholders can work together to develop models that can maximally inform educational policy and practice. Nevertheless, a great deal of statistical and empirical work is required to fully develop the Bayesian approach for the educational sciences.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Box, G., & Tiao, G. (1973). *Bayesian inference in statistical analysis*. New York: Addison-Wesley.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist* , *49*, 997–1003.

Fisher, R. A. (1941/1925). *Statistical methods for research workers* (84th ed.). Edinburgh: Oliver & Boyd.

Fox, J. P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel irt model using gibbs sampling. *Psychometrika*, *66*, 271–288.

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, *100*, 680–700.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis, 2nd edition*. London: Chapman and Hall.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London: Chapman and Hall.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum and Associates.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*, 382–417.

Hoijtink, H., Klugkist, I., & Boelen, P. A. (2008). *Bayesian evaluation of informative hypotheses*. New York: Springer.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

McCandless, L. C., Gustafson, P., & Austin, P. C. (2009). Bayesian propensity score analysis for observational data. *Statistics in Medicine*, *28*, 94–112.

Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, *29A, Part I* , 175–240.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* (70), 41-55.

Appendix B. Tables and Figures

Not included in page count.